Nucleic Acids Research, 1993, Vol. 21, No. 17 3951–3957

Definition of the DNA-binding site repertoire for the *Drosophila* transcription factor SNAIL

Viviane Mauhin, Yves Lutz, Christine Dennefeld and Audrey Alberga*
Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cedex, France

Received June 14, 1993; Accepted July 14, 1993

ABSTRACT

The *Drosophila* gene *snail* (*sna*) which encodes a zinc finger protein is essential for dorsal – ventral pattern formation in the developing embryo. We have defined a repertoire of SNAIL (SNA) binding sites using recombinant SNA proteins to select specific binding sequences from a pool of random sequence nucleotides. The bound sequences which were selected by multiple rounds of gel retardation and amplification by the polymerase chain reaction (PCR) were subsequently cloned and sequenced. The consensus sequence, 5'G/A A/t G/A A CAGGTG C/t A C 3', with a highly conserved core of 6 bases, CAGGTG, shares no significant homology with known binding sequences of other *Drosophila* zinc finger proteins. However, the CAGGTG core is identical to the core motif of aHLH (helix-loop-helix) binding sites. The strongest SNA binding is obtained with sequences containing this core motif whereas reduced binding is seen for sequences with canonical CANNTG HLH motifs. Interestingly, SNA binding is detected in the promoter region of the *snail* gene. Transient expression in co-transfection experiments using a SNA binding element (SBE) linked to a heterologous promoter indicates that SNA has the ability to function as a transcription activator.

INTRODUCTION

A large number of genes involved in the process of *Drosophila* embryogenesis have been identified, and many of them encode DNA-binding proteins that control the transcription of downstream genes (1). While early genetic studies showed that the *sna* gene is essential for normal mesoderm formation (2,3) more recent studies suggest that it is required for the definition of the boundary between mesoderm and neuroectoderm (4,5,6). In addition, our molecular studies showed a complex temporal and spatial *sna* expression in all three germ layers and in particular, in the developing central and peripheral nervous system, which suggested additional but as yet undefined roles for the gene in later embryogenesis and neurogenesis (7). The presence of five Cys_2-His_2 zinc finger motifs in SNA (8) suggests that it is a sequence-specific DNA-binding protein involved in the regulation of RNA transcription. We were interested in obtaining a catalogue of SNA binding sequences as the initial step in the characterization of this putative transcription factor and have used PCR amplification of sequences selected from a pool of random oligonucleotides by the gel mobility shift assay (9,10).

MATERIAL AND METHODS

Binding site selection

Oligonucleotides:double stranded DNA probe. The oligonucleotide synthesized for use in the binding selection, 5'-GATCGGATCCGGATATT(N17)CTCCTGCAGACTCG-ACT-3' (FR18) contained an internal region of 17 degenerate bases flanked by two 17 base sequences containing *Bam* HI (5') and *Pst* I (3') restriction sites. For PCR amplification, the oligonucleotides 5'-AGTCGAGTCTGCAGGAG-3' (FR19) and 5'-GATCGGATCCGGATATT-3' (FR20) served as forward and reverse primers respectively. The oligonucleotides FR18 and FR 19 (100 pmole each) were end-labelled by T4 kinase reactions, annealed and the second strand was synthesized with the addition of the 4dNTP (final concentration of 0.3 mM each) and extension by the Klenow fragment of DNA polymerase. Unincorporated nucleotides were removed by filtration on Sephadex G50 and the labelled double stranded DNA was precipitated by ethanol with added carrier tRNA (5–10μg) and resuspended in double-distilled water (GDW).

Binding reaction and gel retardation. β-GAL-SNA fusion proteins prepared as previously described (7) were pre-incubated with 1μg poly (dI-dC) as a non-specific competitor for 15 min at 0–4°C in a reaction mixture of 20mM Tris–HCl pH 7.5, 0.1mM EDTA, 0.5mM DTT, 1mM $MgCl_2$, 50mM KCl, 0.2mM $ZnSO_4$, 0.2mM PMSF, anti-protease cocktail (10μM TPCK, 10μM TLCK, 15μM pepstatin, 0.1μM leupeptin), 10% glycerol before the addition of 2 pmole of the ^{32}P-DNA to give

* To whom correspondence should be addressed

a final reaction volume of 20μl. The incubation was continued for an additional 20 min. When appropriate, antibodies were added at this stage and the incubation prolonged for another 15 min. The samples were analyzed on a 5% acrylamide gel (39:1 acrylamide/bis) in 0.5× TBE for 1.5−2h at 15−20 volt/cm. The shifted band, corresponding to the protein-DNA complexes, was excised and eluted overnight at 37°C in 0.5mM ammonium acetate, 1mM EDTA, 0.1%SDS, 10mM Tris−HCl pH 7.5. The DNA was extracted with phenol-chloroform and precipitated by ethanol with added carrier tRNA (5−10μg). The recovered DNA was resuspended in appropriate volumes of GDW.

PCR amplification, plasmid construction and sequencing. PCR amplifications were carried out in a final volume of 100μl containing 10 pmole of each primer (one of which was ^{32}P-labelled), 0.2mM of the 4 dNTPs, 10mM Tris−HCl pH8.3, 50mM KCl, 1.5mM $MgCl_2$ and 1μl of Taq polymerase (5u). After an initial denaturation of 3 min at 94°C, amplifications were for 1 min at 94°C, 2 min at 50°C and 3 min at 72°C for 25 cycles in a thermal cycler (Perkin Elmer-Cetus). The amplified DNA was purified on a 12% acrylamide gel, eluted as described above and served as the template for the following round of selection. A total of four selection cycles (binding/gel-shift/PCR) were carried out and the fourth round of amplification was performed with 50 pmole of each non-radioactive primer. The final pool of oligonucleotides was digested with *Bam* HI and *Pst* I, ligated into pBluescript SK$^+$ (Stratagene) and transformed into XL1-Blue cells (Stratagene). The inserted DNA sequences were determined using the Taq Dye-Deoxy Termination Cycle sequencing kit with the M13 primer on a 373A DNA sequencing apparatus (Applied Biosystems).

Analytical gel retardation of selected sequences

Oligonucleotide probes and binding. End labelled double stranded DNA probes were made from derived oligonucleotides of selected sequences flanked by *Bam* HI sites to permit cloning into the *Bam* HI site of pBLCAT8$^+$ (11) in order to construct reporter genes (see below). Binding reactions using various SNA preparations were performed as described above and protein−DNA complexes were resolved on 4% polyacrylamide gels in 0.5×TBE. In competition experiments, cold competing double stranded oligonucleotides were added at the same time as the labelled probes.

Embryonic proteins. Drosophila embryos (0−12h) were pulverized in liquid nitrogen (all subsequent manipulations were performed at 0−4°C) and homogenized with a Dounce homogenizer in 3 vol of extraction buffer (15mM HEPES pH 7.6, 10mM KCl, 5mM $MgCl_2$, 0.1mM EDTA, 0.5mM EGTA, 1mM DTT, 0.35mM sucrose, 2mM PMSF and anti-protease cocktail). Cellular debris was eliminated by filtration through gauze and the homogenate centrifuged at 2000 rpm for 10 min. The soluble fraction (supernatant) was separated from the crude nuclear pellet which was extracted by homogenization and sonication (four 15 sec bursts using a Vibra Cell apparatus, Bioblock) in 15mM HEPES pH 7.6, 200mM KCl, 3mM $MgCl_2$, 0.1mM EDTA, 1mM DTT, 2mM PMSF and anti-protease cocktail. Both extracts were centrifuged at 10,000 rpm for 30 min and dialyzed against 15mM HEPES pH 7.6, 40mM KCl, 5mM $MgCl_2$, 0.1mM EDTA, 1mM DTT, 10% glycerol. Residual material was eliminated by centrifugation and protein concentrations were determined by the Bradford assay (12).

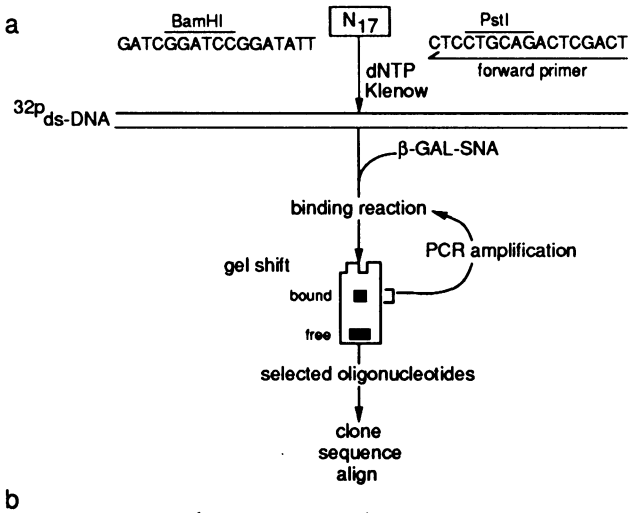
Transfection: whole cell extracts. Schneider line 2 (S2) cells were cultured in Schneider Drosophila medium (Gibco) supplemented with 10% fetal calf serum, COS-1 cells were maintained in Dulbecco's modified medium supplemented with 5% fetal calf serum. The full length cDNA was inserted in the SV40 promoter vector pSG5 (13) and in the actin promoter vector pPac (14). These plasmids were introduced into COS-1 and Schneider S2 lines respectively by standard calcium phosphate transfection techniques (15). The transfected cells were lysed by three freeze-thaw cycles (liquid nitrogen/37°C) in 250 mM Tris−HCl pH 7.6, 5mM DTT, 15% glycerol, 1mM PMSF and anti protease cocktail followed by sonication of four 15 sec bursts (ice/NaCl bath). The extracts were centrifuged at 10,000 rpm for 20 min at 4°C and aliquots of the supernatants were stored at −20°C.

Preparation of monoclonal antibodies. Balb/c mice were injected with β-Gal-SNA fusion proteins either pD12 (aa 1−390) or pB1 (aa 151−390) and hybridomas prepared by standard PEG fusion protocols. The following procedure was used to select the antibodies directed only against SNA. All hybridoma supernatants were screened with a modified immunocytochemical procedure on COS-1 cells transfected with a *sna* expression vector containing the complete coding region in the eucaryotic expression vector pSG5 (see above).The transfected COS-1 cells were fixed in microtiter plates with 2% formaldehyde. The cells were made permeable by treatment with 0.1% Triton X100 (TX), washed with PBS and incubated overnight at 20°C with diluted (2−10 fold) hybridoma supernatants. The cells were washed with PBS/TX and incubated with goat anti-Mouse IgG (F_C fragment) alkaline phosphatase conjugated secondary antibody. The complexes were detected with a solution containing 1mg/ml Fast Red (Pierce), 0.2mg/ml NAMP (Napthol AS-MX phosphate, Fluka) in 100mM Tris−HCl pH 8.2. Positive cultures were subsequently submitted to a differential screening using the different immunogens (pD12, pB1) and β-GAL in ELISA and Western blot analyses. Specific cultures were double cloned on soft agar. The monoclonal antibody used in these studies, 2SN-5H, is directed against residues 151−390.

RESULTS

Selection of binding sequences

To identify DNA sequences that are recognized by the *snail* protein we chose the technique described by Oliphant *et al.* (9) as modified by Mavrothalassitis *et al.* (10). This technique involves the *in vitro* selection of the binding sites from a population of degenerate oligonucleotides by gel retardation and subsequent PCR amplification of the selected sequences. After multiple cycles of selection and amplification, the binding sequences selected by this procedure are determined by cloning and sequencing. β-GAL-SNA fusion proteins were used to select the binding sites from a 51 bp double stranded oligonucleotide containing a random core of 17bp flanked by defined sequences with restriction sites for cloning (Fig.1a). In preliminary experiments three *E.coli* expressed *snail* proteins were tested, one pD, corresponding to the full length protein of 390aa and two truncated forms, pB which includes residues 153−390 and pC, which includes only the putative DNA-binding finger region, residues 245−390. Under our experimental conditions, fusion proteins pD and pB appeared to be less stable than pC, and subsequently only pC was employed for the selection of the binding sites reported in these studies. After 4 sucessive rounds



b

1	aagtagaaa	caggtg cg
2	agcaa	caagtg ccatag
3	ata	caggtg cacagatc
4	a	caggtg tctaacgatc
5	g	caggtg taagctgtg
6	gcaa	caggtg atagggа
7	t	caggtg cacattatgg
8	agaa	caggtg accctta
9	taa	caggtg cagctggc
10	gt	atgatt ccttctcc
11	gaataa	caggtg agtta
12	a	caagtt gagacaacat
13	ggaaaagaaca	caggtg
14	gcacgaaaaa	caggtt g
15	a	caggtg caatatccgc
16		caggtg ctacaagtgcc
17	gaatatggtga	caggtg
18	g	caggtg caataaccct
19	acgaactaag	caggtg t
20	gcgtcgagg	caggtg cc
21	tctatga	caggtg cacc
22	ctaa	caggtg cacaggt
23	ggatga	caggtg catag
24	acaga	cagatg tactca
25	cctgaga	caggtg ttga
26	agcagtgcgg	caggtg t
27	tacgaagca	caggtg ta
28	ccctaaccag	caggtg t
29	g	caggtg cacgggcta
30	ga	cagatg tctccatac
31	tgt	caggtg gaccatca
32		caggtg caggaagatta
33	aa	caggtg tcaccacat
34	caggttgcgga	caggtg
35	tcag	caggtg cacacgc
36	a	caggtg tactcatatg
37	caaa	caggtg ctggaac
38	ga	caggtg ccgacccgta
39	cgga	caagtg cagggag
40	a	caggtg tacgtgatca

c

	-4	-3	-2	-1	C	A	G	G	T	G	+1	+2	+3
A	7	9	13	27	0	39	3	2	0	0	3	18	7
G	7	4	13	8	1	0	37	38	0	37	3	2	7
C	6	5	2	0	39	0	0	0	0	0	19	8	11
T	2	7	1	3	0	1	0	0	40	3	12	4	5
TOTAL	22	25	29	38	40	40	40	40	40	40	37	32	30
FREQUENCY	0.63	0.64	0.89	0.71	0.97	0.97	0.92	0.95	1	0.92	0.83	0.56	0.36
CONSENSUS	G/A	A/t	G/A	A	C	A	G	G	T	G	C/t	A	C

Figure 1. Schematization of SNA binding site selection. (**a**) Labelled random double stranded oligonucleotide was incubated with β-GAL-SNA. The complexes formed were isolated by gel mobility shift, the DNA was extracted, amplified by PCR and served as the template in the following round of selection. (**b**) Alignment of 40 SNA binding sequences recovered after four rounds of selection. Only the nucleotides included in the random region are shown and are aligned around the caggtg motif. Variants of this motif are underlined. (**c**) Consensus SNA binding site derived from the base frequencies of the selected sites.

of selection and amplification, the retarded oligonucleotides were inserted into Bluescript SK$^+$ and 40 clones were sequenced.

An examination of these sequences revealed a CAGGTG motif in 33 of the selected binding sites while the remaining seven contained variants of this motif, the most extreme variant being

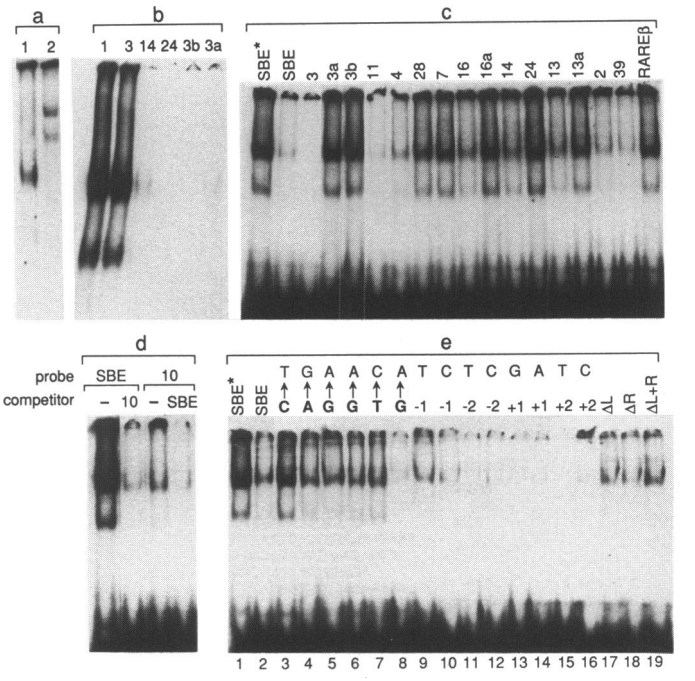


Figure 2. Specificity and binding capacity of selected SNA binding sites. The sequence, ggatgacaggtgcatag, which gave the best fit with the consensus was designated as SBE (SNAIL binding element) and used to determine the characteristics of the selected sequences using standard binding reaction conditions and gel mobility shift assay (see Material and Methods). (**a**) Fusion protein pC (3μg) was incubated with ^{32}P-SBE (300 fmole) without (lane 1) and with (lane2) anti-SNA monoclonal antibody (2SN-5H) added to the binding reaction after formation of the complex. Note the 'super shifted' β-GAL-SNA−DNA complex (slower migrating band) and a faster moving band, due to the complex formed with a proteolysed form of the fusion protein (barely visible in lane 1). (**b**) pC (6μg) was incubated with ^{32}P-target sequences (300 fmole), SBE (lane1), and oligos 3, 14, 24 (Fig. 1b) and 3b, 3a. The last two sequences are point mutations of sequence 3 (see **c** below). (**c**) labelled SBE (20 fmole) was incubated with 0.5μg of fusion protein and competed by unlabelled target DNA (1pmole) which was added to the binding reaction at the same time as labelled SBE. Numbers correspond to the numeration given in Fig. 1b. Modified sequences are indicated by the letter a or b after the number of the sequence. (3a) atccaggtgcaccagatc, (3b) atacaggcgcacagatc, (13a) ggaaaagaacacaggtg minus 3′ flanking Bam H1 site, (16a) caggtgctccaagtgcc minus 5′ flanking *Bam* H1 site. Note that a 100 fold excess of an unrelated DNA binding sequence, RAREβ (retinoic acid β-receptor binding element see Fig. 3c for sequence) does not compete for SBE sites. (**d**) Binding capacity of sequence no. 10. pC (0.5mg) was incubated with 7.2 fmole of either labelled SBE or no. 10 with or without the reciprocal oligonucleotide (50 fold excess) as competitor. On the right, competition of no. 10 for labelled SBE binding and on the left, competition of SBE for labelled no. 10 binding. (**e**) pC (0.5μg) was incubated with ^{32}P-SBE (20 fmole) lane 1 and competed by unlabelled SBE lane 2 and in lanes 3 to 16, unlabelled templates in which single base substitutions were introduced (as shown above) or flanking bases deleted. ΔL, deletion of bases to the left of position −2, ΔR, deletion of bases to the right of position +2, ΔL+R left and right deletions. All competitors were at 1pmole.

no. 10 (see below). The alignment of the 40 independent sequences by this 5′-CAGGTG −3′core (showing only bases included in the degenerate region of each oligonucleotide) is shown in Fig. 1b.

Consensus SNA binding site

An analysis of the base frequencies generated the resulting consensus binding site 5′-G/A A/t G/A A CAGGTG C/t A C-3′ (Fig. 1c). Variation in the core was limited to substitution of an A for the first G (3 cases) or second G (2 cases) and a T for

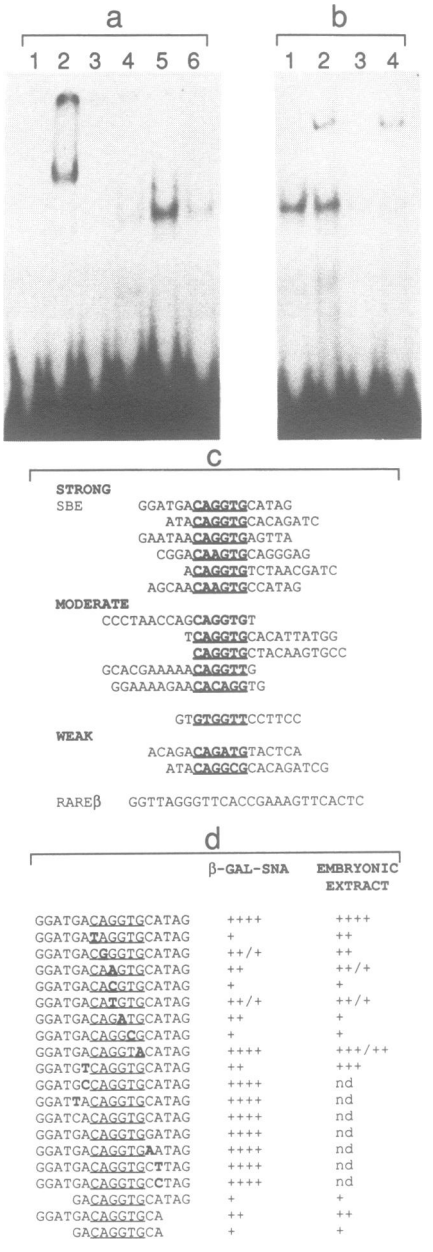


Figure 3. Binding capacity of diverse sequences for β-GAL-SNA and endogenous SNA. (**a**) embryonic protein extracts soluble (2μg), lane 5 and nuclear (2μg), lane 6, were incubated with labelled SBE (50 fmole) and complex formation compared to that obtained with other SNA preparations. *E.coli* fusion protein pC (0.5μg), lane 2, extracts of cells transiently producing SNA, COS-1 (2μg), lane 3, Schneider S2, (0.5μg) lane 4 (see Material and Methods for preparation of proteins), lane 1, SBE-no protein. (**b**) super shift of endogenous SNA−SBE complex by 2SN-5H. Soluble proteins without and with added antibodies, lanes 1 and 2, respectively. Nuclear proteins without and with added antibodies, lanes 3 and 4, respectively. (**c**) classification of sequences according to the observed binding capacities. (**d**) binding efficiencies of mutated SBE for β-GAL-SNA and embryonic proteins (++++, strong, +++, relatively strong, +++/++, moderate to strong, ++, moderate, ++/+, moderate to weak, +, weak, nd, not determined).

the third G (3 cases). In one case both the C and A in the core were changed in the same sequence (no. 10 Fig. 1b). In the flanking sequences, there is a marked preference for purine residues at positions −1, −2 and +2 and for a pyrimidine residue at +1. The sequence no. 23 (Fig. 1b), 5′-GGATGACAGGTG

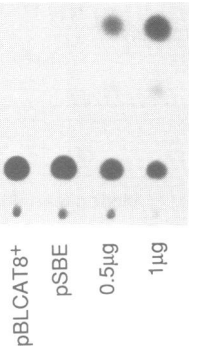


Figure 4. SNA dependent stimulation of reporter gene activity. Schneider cells (10^7) were transfected with 0.5 or 1 μg of *sna* expression vector (pPac-*sna*), 4μg of reporter plasmid (pSBE), 1μg of pACH β-galactosidase expression vector (33) used as an internal control for transfection efficiency, carrier DNA (pSK 11^+) to 20μg (15) and incubated with the precipitate at 25°C for approximately 60 h. Similar transfections without activator were performed with the vector pBLCAT8$^+$ (4μg) and the reporter plasmid, pSBE (4μg). The transfected cells were harvested, washed with TBS (25mM Tris−HCl pH 7.5, 137 mM NaCl, 0.5 mMKCl, 0.7 mM$CaCl_2$, 0.5mM $MgCl_2$, 0.6mM Na_2 HPO_4) and lysed as described in Material and Methods. β-GAL content was determined and CAT asssays were performed on aliquots containing equivalent amounts of b-GAL as described (34).

CATAG-3′, which has the core placed centrally and gave the best fit with the consensus was designated as the SNAIL binding element,SBE, and used as a standard in subsequent experiments.

Selected sequences represent bona fide SNA binding sites

To determine whether or not binding to SBE was specific for SNA, a monoclonal antibody (2SN-5H) directed against SNA (see Material and Methods) was added to the β-GAL-SNA−DNA complex mixture following the binding reaction. This resulted in a super shift of the retarded band, confirming that SNA was responsible for the original band shift (Fig 2a, lane 2). The faster migrating complex (lane 2) contains a proteolysed form of the fusion protein (see also Fig.2b lanes 1 and 2) since this supershift is only seen with 2SN-5H and not with β-Galactosidase antibodies (data not shown).

SNA binding to selected sites

Individual binding sites were tested for their ability to form complexes with SNA. The sequences were chosen so as to be representative of the different patterns obtained and binding efficiency was determined either by direct interactions between the protein and the different radiolabelled DNAs or by competition of the unlabelled target sequences for labelled SBE binding sites. In direct binding experiments, fixed amounts of fusion protein pC were incubated with ^{32}P end-labelled sequences and in competition analysis, SNA was incubated with labelled SBE and a 50 fold excess of unlabelled target DNA. Preliminary experiments had shown that with SBE a 50 fold excess of competitor reduced radiolabelled SBE binding approximately 70−90%(depending on the amount of protein present in the binding reaction).

The results of the gel shift assays which are presented in Figs. 2b and 2c and summarized in Fig. 3c, show a wide range in binding capacities of the different sequences. The strongest binding is observed for sequences with CAGGTG cores with at least one nucleotide on either side (direct binding, Fig. 2b, lanes 1 and 2; competition assays Fig.2C lanes 3 and 11). Relatively

strong binding of SNA is observed with CAAGTG (Fig. 2c lanes 2 and 39) and CAGGTT (Fig. 2c lane 14) core sequences while weak binding is seen with CAGATG cores (Fig. 2b lane 4 and Fig. 2c lane 24). For optimal binding at least one base is required 5′ to the core motif and preferably, this should be a purine residue (Fig. 2c, compare sequence 4 with 7 and 16a). Unexpectedly, the most extreme variant, gt**gtggtt**ccttctcc (see Fig. 1b, no. 10) which maintains only the --GGT- residues of the core motif binds the protein with moderate efficiency and competes for the SBE binding sites (Fig. 2d). Thus despite its limited resemblance to the consensus, this confirms its initial selection (see discussion).

The role of individual SBE bases in binding affinity

A systematic mutagenesis of the central 10 nucleotides of SBE was performed and the binding capacities of the mutated DNAs were evaluated by their ability to compete with labelled SBE for SNA binding (Figs. 2e and 3d). As expected from the results of the previous series of experiments, the most critical nucleotides are those of the CAGGTG core. Sequences in which the C, A, second G and T were substituted by T, G, A and C respectively, are poor competitors for SBE binding, indicating reduced binding capacities (Fig. 2e lanes 3−7). The last nucleotide of the core does not appear to have any significant effect on the binding and an olignucleotide in which this G is exchanged with an A, is an efficient inhibitor of SNA−SBE complex formation (lane 8). The replacement of the first G by an A, results in a moderate reduction of labelled SBE binding (Fig.2e, lane 5 and Fig. 3d), whereas a more pronounced inhibitory effect on the binding is seen when bHLH 'E box' binding motifs (16) are created by substituting this G with a C or a T (Fig. 3d). In the flanking sequences, the purine residue at position −1, is apparently needed for optimal SNA binding as the substitution of pyrimidine residues at this position results in lower binding capacities and these mutated sequences are moderate competitors for SBE sites (Fig. 2e, compare lanes 9 and 10). On the other hand, the residues at positions −2, +1 and +2 do not appear to have much influence on the binding and sequences with mutations at these positions are excellent competitors for SBE binding (Figs. 2e lanes 11−16 and 3d). However, the deletion of the nucleotides to the left of position −2, to the right of position +2 or both simultaneously, results in reduced binding capacity (Figs. 2e lanes.17−19 and 3d). We conclude from the results of these experiments that the specificity of SNA binding is most likely determined by the C-AGGTG core and that the overall affinity is dependent on the flanking sequences.

Endogenous SNA binds to SBE

The *snail* gene is expressed during embryogenesis and since the binding sites were selected with an *E.coli* expressed fusion protein, it was necessary to determine if the same SBE binding pattern was obtained with endogenous SNA. Soluble (cytoplasmic) and nuclear proteins were prepared from *Drosophila* embryos (see Material and Methods) and the binding of these proteins to SBE is presented in Fig. 3a. On gel shifts, the complexes formed between the embryonic proteins and SBE (lanes 5 and 6) migrate the same distance as those formed with the full length protein extracted from cell lines transiently producing SNA, COS-1 (lane 3) and Schneider S2 (lane 4). For identical protein concentrations, more binding activity is observed for the soluble fraction than with the proteins in the nuclear extract. The binding detected must in part be due to SNA as the addition of the anti-SNA monoclonal antibody results in a further

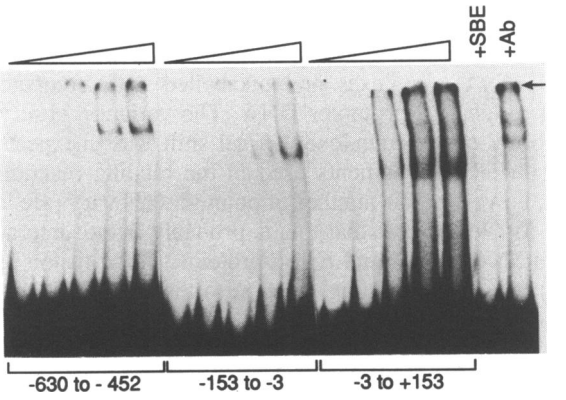


Figure 5. SNA binding to *sna* promoter DNA. Sequences from the *sna* promoter region were labelled by PCR and 15 fmole were incubated with increasing amounts of pC (0,3,6,and 9μg) under standard binding conditions except for an increase in poly (dI.dC) to 2μg and the addition of 2.5μg denatured salmon sperm DNA to the reaction mix. The binding of all target sequence (15 mole) incubated with 3μg pC was competed by SBE (50 fold excess) and gave a supershift with the antibodies, (as shown (arrow) for the region −3 to +153). The faster moving bands (+Ab) are apparently not due to SNA binding as these are already present prior to the addition of the antibody. The coordinates are relative to the transcription initiation site. −630 to −452 contains five putative binding sites: gaaggtgtc, ctcaggttc, ggcagatgt (coding strand), agaggtgtt, ggtggtttc (non coding strand). −153 to −3, three putative sites: ttgttggta, cacacgtgc (coding strand), ggtggttc (non coding strand). −3 to +153, 1 site: ggtggttg (non coding strand).

shift in the mobility of some complexes (Fig. 3b, lanes 1 and 2 (soluble), lanes 3 and 4, nuclear). At present it is not known if the partial shift is due to an excess of protein in the binding reaction, the binding of SNA *in vivo* with other proteins or overlapping recognition sites for multiple proteins. The binding capacity of the endogenous SNA proteins to mutated SBE oligonucleotides is similar to that observed for the *E.coli* expressed fusion protein (Fig. 3d).

Transcription activation by SNA

We have previously reported that *sna* is expressed in two phases during embryogenesis, an early phase which occurs prior to gastrulation and a late phase starting with germ band elongation (8). To determine if during these two phases, SNA could function as an activator and/or a repressor of transcription, *Drosophila* Schneider S2 cells were co-transfected with a *sna* expression vector containing the complete *sna* coding region fused to the *actin 5C* promoter and a reporter plasmid containing a single copy of SBE linked to the chloramphenicol acetyltransferase (CAT) gene (pSBE). As seen on Fig. 4, the addition of 0.5 or 1μg of activator results in a significant increase in transcription activity of the CAT gene of 18 and 32 fold, respectively. At higher concentrations, a decrease in the activation was observed (data not shown) which may be related to the squelching phenomenon observed with other activators in transient transfection experiments (17,18).

SNAIL binding sites in the *sna* promoter

In the course of an analysis of the *snail* promoter we recognized several degenerate copies of the SNA binding core sequence (including gtggtt variants, see above) and we examined the binding of the β-GAL-SNA to several fragments of the *sna* promoter region (Fig. 5). In comparison with SBE binding, the overall binding efficiency is weak and high protein concentrations

were needed to detect binding. However, in all cases, the addition of SNA antibodies gives the expected supershift of the protein–DNA complexes and unlabelled SBE competes for binding to labelled promoter DNA. The variation observed in the mobility of the complexes on gel shift was unexpected, as the length of the fragments used in the binding reactions are similar, however their number of putative sites vary (see legend to Fig. 5). We believe that this is probably due to interactions between SNA–DNA and *E.coli* proteins (in the fusion protein extract) during the binding reaction to these longer templates.

DISCUSSION

The base frequence analysis of 40 selected DNA sequences defined the consensus binding site, G/A A/t G/A A CAGGTG C/t A C, for the SNA *Drosophila* zinc finger protein. This consensus shows no similarity to known recognition sequences of other *Drosophila* Cys_2-His_2 finger proteins (19,20,21). Surprisingly however, the CAGGTG core motif is identical to the binding motif of aHLH proteins such as the immunoglobulin light chain enhancer binding proteins, E12 and E47 (22) and the gene product of *daughterless*, a gene involved in the development of *Drosophila* peripheral nervous system (23). Strong SNA binding sites have intact CAGGTG core motifs, strong to moderate binding sites, CAAGTG or CAGGTT motifs, while weak binding sites have CAGATG core sequences or substitutions which create bHLH binding sites (CACGTG or CATGTG). The selected sequence (no. 10 Fig. 1b) with a GTGGTT core motif has moderate binding capacity, and we note that this motif is found in SV40, and several minimum enhancers (24,25) but also in the *sna* promoter (Fig. 5).

Recently SNA binding sites were identified in the promoter regions of the genes *singleminded(sim)* and *rhomboid* (*rho*) by DNAseI footprinting (26 and 27, respectively). The reported consensus *sim* SNA binding site, TGNNAACAGGTGNT(coding strand), corresponds to a strong site in the present study and the consensus C(A/C)ACTTGC (GCAAGTG (G/T) G) defined in the *rho* study correspond to moderate to strong SNA binding sequences, depending on the flanking sequences.

Genetic studies have suggested that *sna* is required to repress the expression of *sim* and *rho*, genes which are involved in the differentiation of the neuroectoderm (4,5,6). We however show here, that SNA can also function as an activator of transcription and that SBE can confer SNA responsiveness on an heterologous promoter, SNA must therefore possess an activation function. It remains to be seen if the SNA binding sites in *sim* and *rho* promoters represent positive or negative regulatory elements and to determine the parameters involved in the definition of these two types of transcriptional responses. Thus, like many other transcription factors, SNA may be both an activator and repressor of transcription, depending on the nature of the binding element, the promoter context and the cell type (14,28,29,30). It will be interesting to see if this dual role is related to the early and late *sna* expression during embryogenesis and neurogenesis (7).

One regulatory mechanism of 'phased' expression is autoregulation whereby the gene product regulates its own expression in certain cell types (31). A good example of autoregulation is the *Drosophila* segmentation gene *fushi tarazu* (*ftz*) which encodes a homeo-box transcription factor and is expressed in two phases during development (at blastoderm and during neurogenesis). The early *ftz* expression is controlled by a proximal promoter 'the zebra element' and a distal enhancer. While the 'zebra element' is regulated by the products of other segmentation genes, the enhancer is positively regulated by the *ftz* gene product (32).

To determine whether *sna* could also be subjected to autoregulation, we looked for putative SBE sequences in the *sna* promoter and in this study we present evidence for several SNA binding sites in the promoter of its own gene. Future *in vivo* analyses should determine whether or not *sna* is indeed submitted to autoregulation *in vivo* and if so, whether this is positive or negative regulation.

ACKNOWLEDGEMENTS

We are grateful to H.Gronemeyer, G.Richards for helpful discussions, F.Rudert for reagents, S.Vicaire for DNA sequencing, F.Ruffenach, I.Colas and P.Hamann for oligonucleotides syntheses. We also thank M.Acker, B.Heller and all members of the cell culture group for the COS-1 and Schneider S2 cells, V.Schultz and N.Jung for their participation in the preparation of the monoclonal antibodies, C.Werlé and S.Metz for illustrations, B.Boulay and J.M.Lafontaine for the photography. This work was supported by funds from the Institut National de la Santé et de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Centre Hospitalier Universitaire Régional, the Association pour la Recherche sur le Cancer and the Fondation pour la Recherche Médicale. V.M. was supported by a MRES studentship.

REFERENCES

1. Ingham, P.W. (1988) *Nature* **335**, 25–34.
2. Anderson, K.V. and Nüsslein-Volhard, C. (1984) In Malacinski, G.M. and Bryant,S.(eds.), Pattern Formation, MacMillan, New York.
3. Grau, Y., Carteret, C. and Simpson, P. (1984) *Genetics* **108**, 347–360.
4. Nambu, J.R., Franks, R.G., Hu, S. and Crews, S.T. (1990) *Cell* **63**, 63–75.
5. Rao, Y., Vaessin, H., Jan, L Y. and Jan, Y.N. (1991) *Genes Devel.* **5**, 1577–1588.
6. Kosman, D., Ip, Y.T., Levine, M., and Arora, K. (1991) *Science* **254**, 118–122.
7. Alberga, A., Boulay, J.L., Kempe, E., Dennefeld, C. and Haenlin, M. (1991) *Development* **111**, 983–992.
8. Boulay, J.L., Dennefeld, C. and Alberga, A. (1987) *Nature* **330**, 395–398.
9. Oliphant, A.R., Brandl, C.J. and Struhl, K. (1989) *Mol. Cell. Biol.* **9**, 2944–2949.
10. Mavrothalassitis, G., Beal, G. and Papas, T.S. (1990) *DNA Cell Biol.* **9**, 783–788.
11. Klein-Hitpass, L., Ryffel, G.O., Heitlinger, E. and Cato, A.C.B., (1988) *Nucleic Acids Res.* 18, 647–663.
12. Bradford, M.M. (1976) *Anal. Biochem.* **72**, 248–254.
13. Green, S., Issemann, I. and Sheer, E. (1988) *Nucleic Acids Res.* **16**, 369.
14. Krasnow, M.A., Saffman, E.E., Kornfeld, K. and Hogness, D.S. (1989) *Cell* **57**, 1031–1043.
15. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current Protocols in Molecular Biology. John Wiley and Sons, New York.
16. Leonardo, M., Pierce, J.W.and Baltimore, D. (1987) *Science* **236**, 1573–1577.
17. Gill, G. and Ptashne, M. (1987) *Cell* **51**, 121–126.
18. Meyer, M.E., Gronemeyer, H., Turcotte, B., Bocquel, M.T., Tasset, D. and Chambon, P. (1989) *Cell* **57**, 433–442.
19. Stanojevic, D., Hoey, T. and Levine, M. (1989) *Nature* **341**, 331–335.
20. Treisman, J. and Desplan, C. (1989) *Nature* **341**, 335–337.
21. Payre, F. and Vincent, A. (1991) *EMBO J.* **10**, 2533–2541.
22. Murre, C., Schonleber McCaw, P. and Baltimore, D. (1989) *Cell* **56**, 777–783.
23. Caudy, M., Vässin, H., Brand, M., Tuma, R., Jan, L.Y. and Jan, Y.N. (1988) *Cell* **55**, 1061–1067.
24. Locker, J. and Buzard, G. (1990) *DNA Seq.-J DNA Seq. Map.* **1**, 3–11.
25. Faisst, S. and Meyer, S. (1992) *Nucleic Acids Res.* **20**, 3–26.

26. Kasai, Y., Nambu, J.R., Lieberman, P.M. and Crews, S.T. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 3414–3418.
27. Ip,Y.T., Park, R.E., Kosman, D., Bier, E. and Levine, M. (1992) *Genes Dev.* **6**, 1728–1739.
28. Diamond, M.I., Miner, J.N., Yoshinaga, S.K. and Yamamoto, K.R. (1990) *Science* **249**, 1266–1272.
29. Pan, D. and Courey, A.J. (1992) *EMBO J.* **11**, 1837–1842.
30. Jiang, J., Rushlow, C.A., Zhou, Q., Small, S. and Levine, M. (1992) *EMBO J.* **11**, 3147–3145.
31. Serfling, E. (1989) *Trends Genet.* **5**, 131–133.
32. Hiromi, Y. and Gehring, W.J. (1987) *Cell* **50**, 963–974.
33. Thisse, C., Perrin-Schmitt, F., Stoetzel, C. and Thisse, B. (1991) *Cell* **65**, 1191–1201.
34. Bocquel, M.T., Stricker, C., Chambon, P. and Gronemeyer, H. (1989) *Nucleic Acids Res.* **17**, 2581–2595.